期刊学科分类准确性及其对我国英文科技期刊布局的影响

张建东[1, 2]，廖宇[1]，沈哲思[1]*1

1) 中国科学院文献情报中心，北京中关村北四环西路 33 号 100190

2) 中国科学院大学经济与管理学院图书情报与档案管理系，北京中关村北四环西路 33 号 100190

摘 要: **[目的]**研究期刊所属领域及其发文领域的偏差，及其对中国英文科技期刊学科布局的影响; **[方法]**基于论文层级的学科分类体系，应用期刊-研究领域发文相似性，测度期刊所属 WoS 研究领域及其所刊载论文研究领域间的偏差，并对比修正发文偏差后，中国英文科技期刊的实际学科布局和原始布局，最后通过期刊和主题分类案例对比说明本文方法的有效性; **[结果]**在本研究数据集的 290 本中国英文科技期刊中，存在高错误分类倾向的期刊 15 本。由于学科分类的错误，中国英文科技期刊在“ELECTROCHEMISTRY”和“GREEN & SUSTAINABLE SCIENCE & TECHNOLOGY”等 22 个领域的学科布局发生了变动。此外在“GERIATRICS & GERONTOLOGY”和“BUSINESS”等学科可能存在空白布局，在原有的空白学科“PUBLIC ADMINISTRATION”和“IMAGING SCIENCE & PHOTOGRAPHIC TECHNOLOGY”中国英文科技期刊表现可能被低估；**[结论]**基于论文层级分类体系的期刊-领域发文相似性能发现在 WoS 分类中存在错误分类的期刊，而这些错误分类的期刊，会造成某些领域期刊建设的“空白”与与竞争压力低估。

关键词：学科分类体系；科技期刊; 期刊布局

1 背景

学术期刊是科学发展的重要载体。合理的学科布局能够优化学术资源配置，促进科学的持续发展。科学合理的学科分类体系是期刊学科布局的有力保障。当前，学者们多采用 Web of Science 学科分类体系（简称 WoS 学科）进行我国英文科技期刊布局的相关研究。然而，这一学科分类体系在合理性和准确性方面存在问题。仅依赖期刊层级分类体系进行学科布局研究，可能会导致潜在的分析误差，并使基于此的期刊发展策略或科技政策面临一定的风险。因此，系统地识别期刊学科分类体系中的分类误差，以及发现现有学科布局中的潜在空白和重复，对于科技评价具有重要意义。

2 相关研究

2.1 我国英文科技期刊学科布局相关研究

期刊布局分析依托于学科分类体系。目前，我国英文科技期刊的布局分析主要采用期刊层级学科分类体系。张楠等人[1]参照教育学领域 SSCI 期刊 WoS 学科分布情况探讨我国教育学领域期刊学科空白问题，王继红等人[2][3]利用 WoS 学科分析了我国 SCIE 期刊学科分布均衡性和学科空白等，林鹏[4]基于 WoS 学科探讨了我国英文科技期刊学科空白问题，岳东方等人[5]探讨了我国生物学领域 SCI 期刊在 WoS 学科上的分布问题，分析了我国生物学领域的优势学科。由于数据分析的便利和 SCI 的广泛使用，目前关于我国英文科技期刊布局的研究多采用 WoS 学科分类体系，但相关研究结论的可靠性和合理性取决于 WoS 分类体系本身的准确性。WoS 学科分类体系的准确性在科学计量学界被广受质疑，因此基于 WoS 学科分类体系分析我国英文科技布局的结论准确性评估尚缺少相关探讨。

2.2 WoS 学科分类问题相关研究

*作者简介：张建东(ORCID: 0000-0001-8848-7607)，博士生，E-mail:zhangjiandong21@mails.ucas.ac.cn；廖宇，博士后，E-mail: liaoyu@mail.las.ac.cn；

通讯作者：沈哲思（ORCID:0000-0001-8414-7912），博士，副研究员，硕导，E-mail: shenzhs@mail.las.ac.cn

如前所述，以 WoS 学科分类为代表的期刊层级学科分类体系在期刊评价和学科布局等领域已有大量应用。由于期刊层级的论文分类数据易于获取与使用，研究人员基于该体系做了大量研究，如期刊评价、学科布局研究。传统的期刊层级分类体系基于引文分析和专家经验已经有许多研究人员关注到期刊层级学科分类体系的合理性及准确性问题。Rafols 和 Leydesdorff 提出由于类似 WoS 的期刊层级分类的方法旨在帮助信息检索，因此若将其用于科学政策和社会学规律研究时需要谨慎[6]。关于期刊层级学科划分方法的结果准确性方面，Wang 和 Waltman 通过期刊与学科领域的引文关系分析了 WoS 和 Scopus 的分类，发现两个分类划分均可能存在一定问题[7]。Shen 等发现由于期刊的错误分类情况，非物理化学领域期刊参与了物理化学领域的 JCR 排名竞争，从而对物理化学领域期刊造成不利[8]。舒非等发现中国科学引文数据库中的论文在期刊层级的学科划分准确率大约为 50%[9]。Leydesdorff 和 Bornmann 在研究中表明，WoS 分类无法提供足够的分析粒度来在科研评价中进行文献计量标准化[10]。Van Eck 等在研究中得出，即使使用复杂的领域标准化指标，在 WoS 等数目数据库中进行领域间的对比评估时仍会低估某些领域的影响[11]。Zhang 和 Shen 基于学科多样性指标，识别 WoS 的多学科期刊中的错误分类期刊，并发现潜在的多学科期刊[12]。因此在进行期刊布局研究过程中，使用 WoS 学科分类可能会存在一定的分析误差。

尽管目前已有关于论文级别的学科分类体系的研究[13][14][15]，然而，将该粒度分类体系应用于期刊学科布局研究的较少。总之，如何系统地发现期刊学科分类体系中的错误分类，并揭示这些错误分类可能对中国英文期刊布局的影响，这一方面，现有研究关注较少。而相对于期刊层级的分类体系，论文层级的学科分类体系能够将单篇论文映射至其所在研究领域，从而能提供更细的分析粒度，此外，其基于定量算法构建而成，能更客观地揭示科学结构，反映期刊最新的发文学科特征[16]。因此，本研究借助论文粒度的学科分类体系，提出一种基于论文粒度分类体系的期刊-领域发文一致性测度方法，识别和调整错误分类的中国英文科技期刊(以下简称中国期刊)，从而挖掘我国期刊学科布局潜在的学科空白与重复。

3 研究方法和数据

3.1 数据

本研究使用的期刊层级分类体系使用 Web of Science 分类体系(WoS)，涉及期刊为科睿唯安 2023 年发布的期刊引证报告(Journal Citation Reports, JCR)中的 SCI 和 SSCI 期刊，论文层级分类体系使用 InCites 数据库中的 Citation Topics 分类，时间窗为 2021 至 2022 年。最终数据集共包含 12280 种期刊，共 4985059 篇论文，其中中国英文科技期刊 290 本。

Web of Science 分类体系由 254 个学科分类构成，涉及自然科学、社会科学和艺术人文。在该分类体系中每一种期刊被给予一个或多个学科分类。Citation Topics 分类体系使用莱顿聚类算法基于引文网络聚类得到，每一篇论文仅分配一个主题，为宏观-中观-微观三级结构。2021 至 2022 年的发文中，共涉及微观主题 2449 个。

3.2 方法

错误分类领域识别与调整。一个领域中，一种期刊的发文应与这个领域中的其他期刊发文关注的主题尽可能相近。若这种期刊的发文关注的主题与本领域其他期刊差别较大，则这本刊存在一定的分类错误风险，当这个差异超过一定的阈值，这本刊则被视为错误分类。此外，若一种期刊与其本不属于的领域发文主题很接近，这种期刊存在分类遗漏的风险，当当这种相似性超过一定阈值，这种期刊则存在被遗漏的分类。当一种期刊同时满足以上两个条件时，其分类将被从错误分类的领域调整至被遗漏的领域。

本研究中，通过 Citation Topics 论文层次主题分类体系对期刊发文主题、WoS 学科发文主题进行向量表征，首先，使用期刊的发文主题向量V_i表征期刊，并用一个领域中所

有期刊发文向量的均值W_{α}表征领域。然后，，并通过计算期刊向量与所有领域向量的余弦相似度$S_{i\alpha}$获得期刊与不同领域的发文一致性。期刊领域、学科领域和相似度计算如下。

期刊发文领域表征。首先，统计期刊在 Citation Topics 微观主题上的发文数量，得到期刊 *i* 的发文向量V_i，该向量中领域v_t表示 Citation Topics 的微观主题 *t*。

$$V_i = [v_1, v_2, \ldots, v_t]$$

WoS 领域表征。WoS 的期刊层级学科分类框架中，每一个分类是一系列期刊的集合，因此，本研究使用领域内，期刊发文的均值表示每一个 WoS 领域W。其中V_i至V_n是领域α中的 *n* 种期刊的发文向量，领域向量W_{α}为这些期刊的发文向量均值。

$$W_{\alpha} = \frac{\sum_{i=1}^{n} V_i}{n}$$

期刊-领域发文相似性计算。用 Cosine 相似度计算每种期刊和每个 WoS 领域的相似性，即 S。为了排除期刊自身发文对领域向量的影响，在计算一种期刊与自身所属领域相似性时，所属领域的发文向量排除了该期刊的发文。

$$S_{i\alpha} = Cosine(V_i, W_{\alpha})$$

期刊发文偏离程度。y 是节点与其最相近领域的相似度，x 是节点与其所属 WoS 领域的相似度。

$$D = \frac{|y - x|}{\sqrt{2}}$$

接下来，统计每一种期刊与自身所属领域一致性以及其最相近领域的相似度的分布，基于期刊在分布图中距对角线的距离识别分类错误期刊。若该期刊的距对角线距离超过 95%的期刊，则该期刊的分类被视为错误分类。最后用该期刊的最相近领域代替该期刊原有的所属领域。

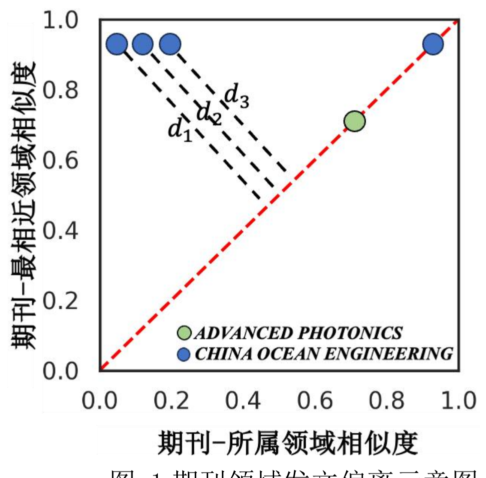


图 1 期刊领域发文偏离示意图

以图 1 的期刊 ADVANCED PHOTONICS 和 CHINA OCEAN ENGINEERING 为例，表 1 展示了每种期刊和期刊与领域相似度表现。

表 1 示例期刊及其领域相似度

期刊	所属领域 及发文相似性 – x 轴	最相近领域 及发文相似性 – y 轴	对角线 距离
ADVANCED PHOTONICS	0.71 (OPTICS)	0.71 (OPTICS)	0
CHINA OCEAN ENGINEERING	0.06 (WATER RESOURCES)	0.94 (ENGINEERING, OCEAN)	0.62
	0.10 (ENGINEERING, MECHANICAL)	0.94 (ENGINEERING, OCEAN)	0.59
	0.18 (ENGINEERING, CIVIL)	0.94 (ENGINEERING, OCEAN)	0.53
	0.94 (ENGINEERING, OCEAN)	0.94 (ENGINEERING, OCEAN)	0

期刊 *ADVANCED PHOTONICS* 所属领域同时也是与其发文相似度最接近的领域 “OPTICS”，因此，该期刊位于在对角线上。而多分类期刊 *CHINA OCEAN ENGINEERING* 属于“ENGINEERING, MECHANICAL”、“WATER RESOURCES”、“ENGINEERING, CIVIL”和“ENGINEERING, OCEAN”四个领域。与该期刊的发文最相近领域同时也是其所属领域“ENGINEERING, OCEAN”。而当该期刊属于“ENGINEERING, MECHANICAL”、“WATER RESOURCES”和“ENGINEERING, CIVIL”时，对角线的距离均超过了 95%的样本(D 小于 0.36)，因此，以上三个领域为该期刊的错误分类，最后该期刊该被划入其最相近领域“ENGINEERING, OCEAN”，而从最终结果看，看期刊从一本四分类的期刊，变成了仅属于“ENGINEERING, OCEAN”的期刊。

4 结果与分析

4.1 错误分类期刊识别

所有 WoS 期刊的发文主题与领域发文主题相似度如图 2 所示。从图 2.a 中可见，一部分期刊的所属领域就是该期刊的最相近领域，这部分期刊分布于该图的对角线上,并且这部分期刊与领域的相似度值主要分布于(0.3,1)区间。同时，也有很多期刊并不位于对角线上，即其发文主题最相近领域，并不是其所属的 WoS 领域。

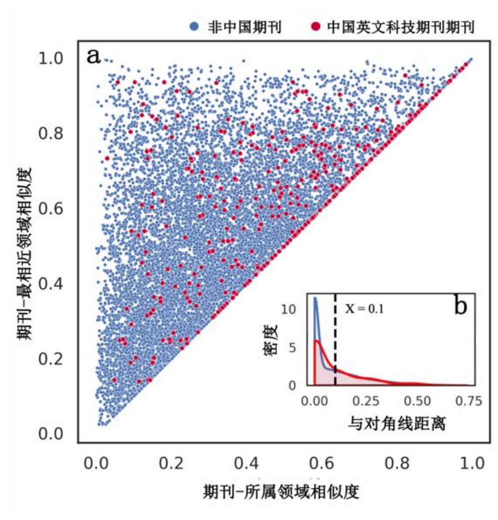


图 2 期刊-领域相似度分布

离对角线越近但是又不位于对角线上的期刊，说明其虽然未被分到发文最相近的的领域，但是该领域与其所属 WoS 领域的发文越相近。而离对角线越远的期刊，说明与其发文主题最相似的领域与其所属的 WoS 领域差别越大。图 2.b 展示了每种期刊与对角线距离的分布。从图中可见，大部分期刊与对角线距离均聚集于(0,0.1)之间，说明大部分期刊的发文主题与其所属 WoS 领域相对一致。其中的非中国英文期刊的对角线距离更为集中，密度最大值为 11.5，中国英文期刊的发文主题与其所属的 WoS 领域发文距离相对更为分散，密度最大值为 5.9。

表 2 不同筛选条件下错误分类期刊数

对角线距离	百分比	实例数(百分比)		期刊数(百分比)	
		中国期刊	非中国期刊	中国期刊	非中国期刊
0.36	95%	20(7%)	963(8%)	15(5%)	794(7%)
0.28	90%	44(15%)	1922(16%)	34(12%)	1553(13%)
0.18	80%	68(23%)	2881(24%)	56(19%)	2283(19%)
0.15	75%	82(28%)	3850(32%)	69(24%)	3010(25%)

当错误分类的阈值选择为 95%时，对角线距离大于 0.36 的期刊被识别为错误分类期刊(表 2)，共有实例数 983 个，即图 2.a 中展示的节点，其中中国期刊实例数 20 个，占所有中国期刊实例数的 7%。由于期刊存在多分类情况，因此其中共涉及期刊 809 本，其中中国错误分类期刊为 15 本，占所有中国期刊的 5%。同时也可以注意到，非中国期刊的错误分类情况略高于中国期刊，说明我们在进行期刊国际对标的时候也要充分注意期刊学科分类错误的情况。

4.2 中国英文科技期刊布局变化

错误分类的 15 本中国英文科技期刊共涉及 WoS 领域 29 个，其中共 22 个领域发生了期刊数量变化，7 个领域新增期刊数与减少期刊数一致。表 3 展示了 22 个发生布局变化的领域及其期刊变化情况。其中，“ELECTROCHEMISTRY”领域新增的期刊最多，为两本，分别为 *JOURNAL OF ENERGY CHEMISTRY* 和 *SUSMAT*。领域“GREEN & SUSTAINABLE SCIENCE & TECHNOLOGY”的期刊数变化最大，从 4 本减少到 1 本。该领域减少的期刊为 *GREEN ENERGY & ENVIRONMENT*，*JOURNAL OF SUSTAINABLE CEMENT-BASED MATERIALS* 和 *SUSMAT*。

表 3 各领域中国英文科技期刊数变化

变化数	领域	原有期刊数	现有期刊数
2	ELECTROCHEMISTRY	1	3
1	MEDICINE, GENERAL & INTERNAL	4	5
	ENGINEERING, GEOLOGICAL	2	3
	OPERATIONS RESEARCH & MANAGEMENT SCIENCE	2	3
	PUBLIC ADMINISTRATION	**0**	**1**
	IMAGING SCIENCE & PHOTOGRAPHIC TECHNOLOGY	**0**	**1**
-1	GEOSCIENCES, MULTIDISCIPLINARY	8	7
	AGRONOMY	3	2
	PLANT SCIENCES	13	12
	CHEMISTRY, APPLIED	4	3

	SOCIAL SCIENCES, MATHEMATICAL METHODS	2	1
	AREA STUDIES	2	1
	UROLOGY & NEPHROLOGY	2	1
	ENVIRONMENTAL SCIENCES	13	12
	MATERIALS SCIENCE, MULTIDISCIPLINARY	26	25
	ENGINEERING, MECHANICAL	8	7
	WATER RESOURCES	4	3
	GERIATRICS & GERONTOLOGY	**1**	**0**
	BUSINESS	**1**	**0**
-2	ENGINEERING, CIVIL	5	3
	ENERGY & FUELS	9	7
-3	GREEN & SUSTAINABLE SCIENCE & TECHNOLOGY	4	1

此外，原有的两个中国英文科技期刊的空白领域，“PUBLIC ADMINISTRATION”和“IMAGING SCIENCE & PHOTOGRAPHIC TECHNOLOGY”有新增期刊，这两个领域均新增了一种期刊。即从 WoS 分类的角度，虽然中国英文科技期刊虽未涉及这两个领域，但是从实际发文分析，中国英文科技期刊已在相关主题有布局。而领域“GERIATRICS & GERONTOLOGY”和“BUSINESS”成为了领域更新后的“空白”领域。

4.3 期刊分类错误案例分析-以期刊 *CHINA OCEAN ENGINEERING* 为例

图 3 是期刊 *CHINA OCEAN ENGINEERING*，以及其四个所属的 WoS 类别“ENGINEERING, MECHANICAL”、“WATER RESOURCES”、“ENGINEERING, CIVIL”和“ENGINEERING, OCEAN”的发文主题。根据图 1.a 的结果，该期刊分属于这四个类别时，与对角线的距离分别为 0.59，0.62，0.54 和 0。根据这一距离，“ENGINEERING, MECHANICAL”、“WATER RESOURCES”和“ENGINEERING, CIVIL”被视为该期刊的错误分类。

图 3 展示了该期刊和每一个领域的前 5 个发文主题，y 轴展示了一种期刊或一个领域的论文所属的研究领域。期刊 *CHINA OCEAN ENGINEERING* 主要涉及的主题是 Ocean Dynamics, Geotechnical Engineering 和 Modelling & Simulation，其中 Ocean Dynamics 占比最高，占整个期刊发文的 45.6%。该期刊关注的主要主题与领域“ENGINEERING, OCEAN”关注的主题基本一致，该领域主要关注 Ocean Dynamics, Moddeling & Simulation 和 Automation & Control Systems，分别占所有发文主题的 30.3%，11.9%和 7.5%。

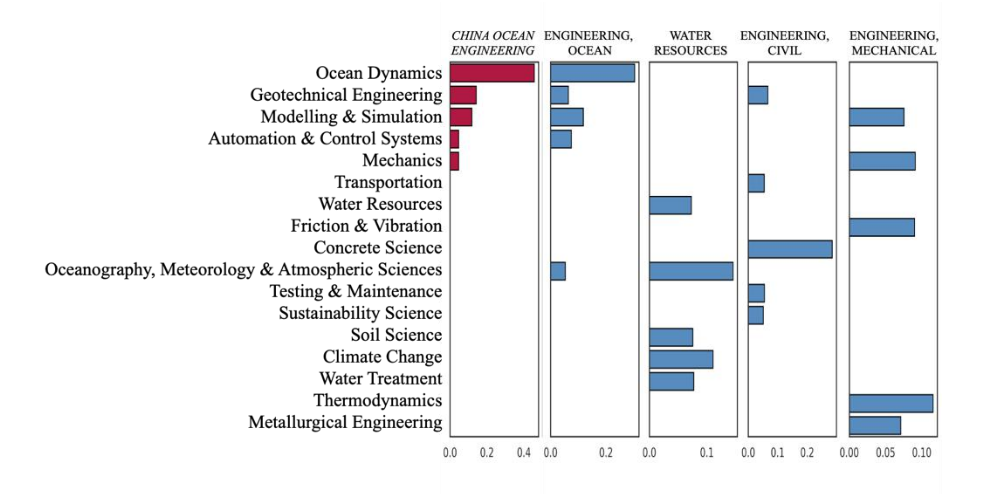


图 3 期刊及所属 WoS 领域发文主题分布——期刊 *CHINA OCEAN ENGINEERING*

然而，对于领域“WATER RESOURCES”、“ENGINEERING, CIVIL”和“ENGINEERING, MECHANICAL”，它们关注的发文主题与 *CHINA OCEAN ENGINEERING* 重叠度却较低。属于“WATER RESOURCES”领域的期刊，主要关注 Oceanography, Meteorology & Atmospheric Sciences(14.5%)和 Climate Change(11.0%)等相关的研究问题；而属于“ENGINEERING, CIVIL”领域的期刊，主要关注 Concrete Science(28.5%)等相关的研究问题，“ENGINEERING, MECHANICAL”领域的期刊主要关注 Thermodynamics(11.6%), Mechanics(9.1%), Friction & Vibration(9.0%)等主题的研究问题。虽然期刊 *CHINA OCEAN ENGINEERING* 与领域“ENGINEERING, CIVIL”和“ENGINEERING, MECHANICAL”有一定的领域重叠，但是该期刊刊载的论文关注的主要研究问题，并不是这两个领域刊载的论文主要关注的。

4.3 期刊分类错误案例分析-以领域 ELECTROCHEMISTRY 为例

图 4 展示了 ELECTROCHEMISTRY 领域以及期刊 *ELECTROCHEMICAL ENERGY REVIEWS*，*JOURNAL OF ENERGY CHEMISTRY* 和 *SUSMAT* 所刊载论文主要关注的领域。在原有分类中期刊 *ELECTROCHEMICAL ENERGY REVIEWS* 属于 ELECTROCHEMISTRY，其刊载的论文主要关注 Electrochemistry 领域的问题，占所有刊载文章的 80.9%。

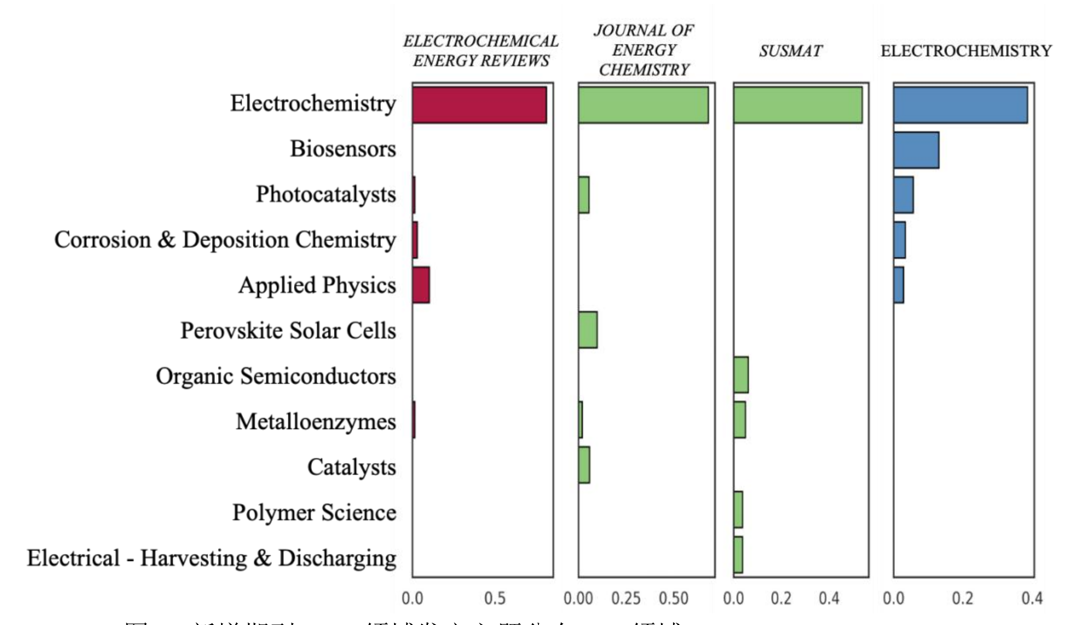


图 4 新增期刊-WoS 领域发文主题分布——领域 ELECTROCHEMISTRY

期刊 *JOURNAL OF ENERGY CHEMISTRY* 和 *SUSMAT* 是本研究中识别为应被分入领域 ELECTROCHEMISTRY 的期刊，其在图 2 中与对角线距离分别为 0.50 和 0.44，均大于 95%的实例(0.36)。而这两种期刊刊载论文所关注的研究问题同样主要关注 Electrochemistry 相关的问题，占主要发文的 68.1%和 53.8%，均与领域 ELECTROCHEMISTRY 刊载论文主要关注的主题相近。同样，该结果与基于对角线识别的结果一致。

5 结论与讨论

本研究针对期刊分类准确性及其对中国英文期刊学科布局的影响问题，基于论文层级分类体系 Citation Topics，表征期刊以及 WoS 领域的发文主题向量，并根据该向量测度期刊与 WoS 领域间的发文一致性，然后借助该一致性的分布识别并纠正错误分类期刊，从而估计中国英文科技期刊的实际布局，最后借助期刊以及领域的案例证明本研究方法的有效性。本研究发现我国的 290 本英文科技期刊中，有 15 本存在错误分类的情况，使 22 个领域的学科

布局与原始布局产生差异，其中差异最大的是“GREEN & SUSTAINABLE SCIENCE & TECHNOLOGY”，其减少了三种期刊。此外，“GERIATRICS & GERONTOLOGY”和“BUSINESS”学科存在潜在的空白布局，而在原有的空白领域“PUBLIC ADMINISTRATION”和“IMAGING SCIENCE & PHOTOGRAPHIC TECHNOLOGY”学科的表现可能被低估。

通过对我国英文科技期刊学科分布的分析，能够掌握我国当前科技期刊发展状态，制定合理的发展战略。论文层级的学科分类体系能够更加精准地反映期刊出版内容的所属学科，以此为依据的期刊学科分布更加接近真实图景，能在多个应用场景中发挥积极作用。对于科技期刊政策制定和实施者来说，本研究利于准确评估各学科的期刊发展状态，有助于制定合理的科技政策，实现更高效的资源分配，增强宏观调控精准性，最终促进学科发展。对于科技期刊运营和编辑人员，可利用本研究结果在选择对标期刊时突破传统数据库学科划分，识别出更多隐性的对标期刊，实现精准对标；在选题和组稿时，也能更加定位于聚焦性的话题和最适配的作者和读者群体。对于科技期刊评价和科学计量学者，本研究结论进一步揭示了现有学科分类体系存在的缺陷和问题，呼吁相关研究者构建更加合理的期刊学科分类体系和推动数据库商尽快优化期刊学科分类，以实现更加准确的期刊影响力评价。

参考文献

[1] 张楠,黄新.教育学期刊结构布局的国内外比较研究——基于 SSCI 和 CSSCI 的分析[J].中国科技期刊研究,2019,30(05):551-558.

[2] 王继红,刘灿,邓群等.我国 SCIE 收录期刊学科分布及建议[J].编辑学报,2015,27(06):576-579. DOI: 10.16811/j.cnki.1001-4314.2015.06.029.

[3] 王继红,刘灿,邓群等.建设 SCIE 空白学科期刊　提升科技期刊国际影响力[J].中国科技期刊研究,2015,26(12):1336-1343.

[4] 林鹏.关于建设世界一流科技期刊的思考与探索[J].中国出版,2020(09):15-20.

[5] 岳东方,宋婷,朱锋荣等.我国生命科学领域优势学科与办刊需求分析[J].学报编辑论丛,2018(00):7-22.

[6] Rafols I, Leydesdorff L. Content-based and Algorithmic Classifications of Journals: Perspectives on the Dynamics of Scientific Communication and Indexer Effects[J]. Journal of the American Society for Information Science & Technology, 2010, 60(9):1823-1835.

[7] Wang Q, Waltman L. Large-Scale Analysis of the Accuracy of the Journal Classification Systems of Web of Science and Scopus[J]. arXiv e-prints, 2015.

[8] Shen Z, Zhang J, Zeng A. Under-representativeness of Physical Chemistry Journals. ChemRxiv. 2024; doi:10.26434/chemrxiv-2024-xj25g

[9] Shu F, Julien C A, Zhang L, et al. Comparing journal and paper level classifications of science[J]. Journal of Informetrics, 2019, 13(1):202-225.

[10] Leydesdorff L, Bornmann L. The operationalization of "fields" as WoS subject categories (WCs) in evaluative bibliometrics: The cases of "library and information science" and "science & technology studies"[J]. Journal of the Association for Information Science and Technology, 2014, 67(3).

[11] Eck N V, Waltman L, Raan A V, et al. Citation Analysis May Severely Underestimate the Impact of Clinical Research as Compared to Basic Research[J]. Plos One, 2012, 8(4):e62395.

[12] Zhang J, Shen Z. Analyzing journal category assignment using a paper-level classification system: multidisciplinary sciences journals[J]. Scientometrics, 2024: 1-16.

[13] Waltman L, Van Eck N J. A new methodology for constructing a publication‐level classification system of science[J]. Journal of the American Society for Information Science and Technology, 2012, 63(12): 2378-2392.

[14] Szomszor M, Adams J, Pendlebury D A, et al. Data categorization: Understanding choices and outcomes[R]. The Global Research Report from the Institute for Scientific Information, 2021.

[15] Boyack K W, Newman D, Duhon R J, et al. Clustering more than two million biomedical publications: Comparing the accuracies of nine text-based similarity approaches[J]. PloS one, 2011, 6(3): e18029.

[16] 廖宇,沈哲思,李立等.论文层次分类体系在期刊学科布局中的应用[J].中国科技期刊研究,2022,33(04):513-520.

致谢：感谢孟平提供有价值的讨论。

英文摘要

The accuracy of journal subject classification and its impact on the layout of Chinese English scientific journals

Abstract: [Purpose] To study the deviation between the journal's field and the field of publication, and its impact on the discipline layout of Chinese English scientific journals; **[Methods]** Based on the paper-level discipline classification system, the journal-research field publication similarity was applied to measure the deviation between the journal's WoS research field and the research field of the papers published, and the actual discipline layout of Chinese English scientific journals was compared with the original layout after correcting the publication deviation. Finally, the effectiveness of this method was demonstrated by comparing the journal and subject classification cases; **[Results]** Among the 290 Chinese English scientific journals in the dataset of this study, 15 journals had a high tendency of misclassification. Due to the error in discipline classification, the discipline layout of Chinese English scientific journals in 22 fields such as "ELECTROCHEMISTRY" and "GREEN & SUSTAINABLE SCIENCE & TECHNOLOGY" has changed. In addition, there may be blank layouts in disciplines such as "GERIATRICS & GERONTOLOGY" and "BUSINESS", and the performance of Chinese English scientific journals in the original blank disciplines "PUBLIC ADMINISTRATION" and "IMAGING SCIENCE & PHOTOGRAPHIC TECHNOLOGY" may be underestimated; **[Conclusion]** Based on the journal-field publication similarity performance of the paper hierarchical classification system, it can be found that there are misclassified journals in the WoS classification, and these misclassified journals will cause "blanks" in the construction of journals in certain fields and underestimate the competitive pressure.

Keywords: discipline classification system; scientific journals; journal layout